Fall 2010

CS 705: Introduction to Data Mining

Guozhu Dong
Wright State University - Main Campus, guozhu.dong@wright.edu

CS 705 Introduction to Data Mining
Fall Quarter, 2010

Description: Data mining is concerned with the extraction of novel and useful knowledge from large amounts of data. This course introduces and studies the fundamental concepts, issues, tasks and techniques of data mining. Topics include data preparation and feature selection, association rules, classification, clustering, evaluation and validation, scalability, mining of spatial/text/sequence/graph/time-series etc data, privacy, data mining applications, and other topics of interest. 3 hours lecture, 2 hours lab.

Prerequisite: CS 605 (Introduction to Database Systems), or CS 609 (Introduction to AI), or equivalent, or consent of the instructor. Implicitly, CS 600 (Data Structures) is also required.

Instructor: Dr. Guozhu Dong. 383 Joshi.

Phone & Email: (937)-775-5066, guozhu.dong@wright.edu

Class details: 4:10-5:25 MW, MM 202

Office hours: 2:30 - 3:25, MW. Use e-mail for short questions. The instructor will also be available (in the class room) for a short period after the classes.

Text Book: Data Mining: Concepts and Techniques, 2nd edition. J. Han and M. Kamber. Morgan Kaufmann.

Reference texts: Introduction to Data Mining, Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Addison Wesley.

Data Mining: Practical Machine Learning Tools and Techniques, I.H. Witten and E. Frank, Morgan Kaufmann.

Principles of Data Mining. D. Hand, H. Mannila, and P. Smith. MIT Press.

Resources: Slides will be available at WebCT.

Students may find this webpage useful: http://www.kdnuggets.com/, especially its pointers to datasets.

Many Java programs for data mining are available at www.cs.waikato.ac.nz/ml/weka, which you may want to install and experiment with.

Grading: Homeworks: 10%, Midterm: 25%; Final: 35%; Projects 30%.

Final grade: A=[90,100], B=[80,90), C=[70,80), D=[60,70), F=[0,60).

The instructor may curve the final grades in such a way that they deviate from these standards at his discretion.

The projects require extensive programming. Submissions that do not compile or that do not address project requirements will receive zero or very low marks.

Handouts: Handouts, and other course material will be distributed in class. It is the students' responsibility to collect them.

Important dates:

4:10 – 5:25, Wednesday, 10/6, in-class midterm exam.

5:45-7:45, Monday, 11/15, final exam.

There will be no class on Wednesday September 29, due to conference trip by the professor.